Mon. Not. R. Astron. Soc. **000**, 000–000 (0000) Printed 4 January 2011 (MN LATEX style file v2.2)

A redshift distortion free correlation function at third order in the nonlinear regime

Kelai Meng[1], Jun Pan[1,2⋆], István Szapudi[2] and Longlong Feng[1]
[1]*Purple Mountain Observatory, 2 West Beijing Rd., Nanjing 210008, P. R. China*
[2]*Institute for Astronomy, University of Hawaii, 2680 Woodlawn Dr., HI 96822, USA*

4 January 2011

ABSTRACT

The zeroth-order component of the cosine expansion of the projected three-point correlation function is proposed for clustering analysis of cosmic large scale structure. These functions are third order statistics but can be measured similarly to the projected two-point correlations. Numerical experiments with N-body simulations indicate that the advocated statistics are redshift distortion free within 10% in the non-linear regime on scales $\sim 0.2-10h^{-1}$Mpc. Halo model prediction of the zeroth-order component of the projected three-point correlation function agrees with simulations within $\sim 10\%$. This lays the ground work for using these functions to perform joint analyses with the projected two-point correlation functions, exploring galaxy clustering properties in the framework of the halo model and relevant extensions.

Key words: cosmology: theory — dark matter — large scale structure of Universe — methods: statistical

1 INTRODUCTION

Observed large scale structure in the Universe is generally conjectured to arise from Gaussian initial condition or nearly so; the rather high level non-Gaussianity at present is due to the action of gravitational force and gas physics. The three-point correlation function (3PCF) is of the lowest order among correlation functions capable of probing such non-Gaussianity. With the recent increase of interest and the corresponding attempts to extract more information about structure formation processes and primordial non-Gaussianity from fine clustering patterns of galaxies, the 3PCF (or its counterpart in Fourier space, bispectrum) has attracted much attention in recent years (e.g Kayo et al. 2004; Nichol 2006; Smith et al. 2008; Jeong & Komatsu 2009; Sefusatti 2009).

However, 3PCF is well known for its low return of investment compared with the two-point correlation function (2PCF). One major obstacle hindering the interpretation and consequently the application of 3PCF is the redshift distortion induced by the peculiar velocities of galaxies. Although effects of redshift distortion on 2PCF (or power spectrum) are not yet well understood analytically (e.g. Scoccimarro 2004), approximations by incorporating pairwise velocity distribution have been proposed, validated and applied successfully to statistical analyses (Peebles 1980; Davis & Peebles 1983; White 2001; Seljak 2001; Kang et al. 2002; Tinker 2007; Smith et al. 2008). In the case of 3PCF (or bispectrum) analogous approach would involve higher order statistics of peculiar velocities. The complicated entanglement of redshift distortions with nonlinear gravitational dynamics and nonlinear biasing renders theoretical prediction extremely difficult in configuration space. In Fourier space and with the distant observer approximation, prediction of the bispectrum in redshift space in various perturbative and empirical schemes has been moderately successful, although none have been able to show satisfactory agreement with simulations (Matsubara & Suto 1994; Hivon et al. 1995; Verde et al. 1998; Scoccimarro et al. 1999). The mostly accurate model to date appears to be the work of Smith et al. (2008), a halo model extension implemented with higher order perturbation theory.

One can eliminate the complexity of redshift distortion with projection of the correlation functions upon the plane perpendicular to the line-of-sight (LOS). Projected correlation functions are obtained by integrating over the anisotropic correlation functions along LOS, which effectively removes redshift distortions if the conservation of total number of galaxy pairs and triplets along LOS can be satisfied. Since thickness of a realistic sample is finite, galaxies near radial edges could enter or leave the sample space by their apparent movement due to peculiar velocity, such conservation is only approximately achieved if the sample is shallow, or redshifts are photometric. Violation of the conservation condition may bring non-negligible systematical bias on large scales (Nock et al. 2010). Nevertheless, this is not a problem for most modern spectroscopic galaxy samples, and the bias actually can be minimized by careful design of estimation methodology.

In comparison with the projected 2PCF that has been widely used to investigate clustering dependence on galaxy intrinsic properties, evolution history and environment and to distinguish cosmological models (e.g. Hawkins et al. 2003; Zheng & Weinberg

⋆ jpan@pmo.ac.cn

2007; Baldauf et al. 2010; Zehavi et al. 2010), exploration and application of the projected 3PCF has been limited in the literature (Jing & Börner 1998, 2004; Zheng 2004; McBride et al. 2010). Lack of accurate theoretical models of 3PCF prevents proper interpretation of measurements. Scoccimarro & Couchman (2001) offered a phenomenological model based on hyper-extended perturbation theory for the bispectrum in the nonlinear regime. Their fitting formula is accurate on smaller scales but in the weakly and mildly nonlinear regimes it is improved upon by the empirical model of Pan et al. (2007). Both fail on very small scales, and neither can capture the signal of baryonic oscillation in bispectrum appropriately (Sefusatti et al. 2010). The approach of halo model appears more promising, as it can reproduce most measurements in simulations for the bispectrum (e.g. Ma & Fry 2000a,b; Scoccimarro et al. 2001; Smith et al. 2006, 2008) and the 3PCF in configuration space (e.g. Takada & Jain 2003; Wang et al. 2004; Fosalba et al. 2005). In spite of disagreement with simulations for some configurations of 3PCF, the halo model is still more attractive than the phenomenological models for its clean and physically motivated parametrization to galaxy biasing through e.g. the machinery of the halo occupation distribution (HOD, Berlind & Weinberg 2002).

Another reason for the scarce exploration of projected 3PCF is the complexity of estimation. Computational requirement of 3PCF is demanding for currently available computers when millions of points are typical. The additional task of decomposing the separations among three points for projected 3PCF adds to the CPU load. Furthermore, the 3PCF is already more prone to Poisson noise than the 2PCF, and typical bin width of scales for projected 3PCF is even smaller than for the normal 3PCF. In order to suppress discreteness effects for a reliable estimation, a high number density of points in the sample is crucial, but often unrealistic for real surveys.

By analogy to the monople of 3PCF advocated by Pan & Szapudi (2005a,b), we show that a third-order statistical function similar to the angular average of the projected 3PCF is redshift distortion free and relatively easy to estimate and model theoretically. In the next section, the definitions, and relation with 3PCF together with estimation algorithm is described. Section 3 presents numerical properties of the new statistical measure while in section 4 we demonstrate the consistency of halo models to simulations of the new function. Summary and discussion are in the last section.

2 PROJECTED THREE-POINT CORRELATION FUNCTION AND ITS ZEROTH-ORDER COMPONENT

Let $\mathbf{r} = \mathbf{x}_2 - \mathbf{x}_1$ be the vector pointing to a point at position $\mathbf{x}_2$ from point at $\mathbf{x}_1$, the vector can be decomposed to two components, separation along the line-of-sight (LOS) $\pi = r\mu$ with μ being the cosine of the angle between the LOS and $\mathbf{r}$, and separation perpendicular to LOS $\sigma = (r^2 - \pi^2)^{1/2}$, then we have the anisotropic 2PCF $\xi(\sigma, \pi)$ and so the projected 2PCF

$$\begin{aligned}\Xi(\sigma) \equiv& \int_{-\infty}^{+\infty} \xi(\sigma,\pi)\mathrm{d}\pi \\ =&2\int_{\sigma}^{+\infty} \frac{r\xi(r)\mathrm{d}r}{\sqrt{r^2-\sigma^2}} = 2\int_{\sigma}^{+\infty} \frac{s\xi(s)\mathrm{d}s}{\sqrt{s^2-\sigma^2}} ,\end{aligned} \tag{1}$$

where $\mathbf{s}$ is the separation vector between two points measured in redshift space and the last step comes from conservation of total number of pairs along LOS. Inversion of $\Xi(\sigma)$ could directly render 2PCF $\xi(r)$ although inversion of such Abel integration is unstable mathematically (Davis & Peebles 1983).

Similarly, giving three points at $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$, 3PCF $\zeta(r_1, r_2, r_3)$ is of the the triangle configuration with three separations $\mathbf{r}_1 = \mathbf{x}_2 - \mathbf{x}_1 = (\sigma_1, \pi_1)$, $\mathbf{r}_2 = \mathbf{x}_3 - \mathbf{x}_2 = (\sigma_2, \pi_2)$ and $\mathbf{r}_3 = \mathbf{x}_1 - \mathbf{x}_3 = (\sigma_3, \pi_3)$, decomposition of the three separations bring up anisotropic 3PCF $\zeta(\sigma_{1,2,3}; \pi_{1,2,3})$ with $\sum \pi_{1,2,3} = 0$, and the projected 3PCF is just defined as (Jing & Börner 1998, 2004; Zheng 2004)

$$\begin{aligned}Z(\sigma_1,\sigma_2,\sigma_3) \equiv& \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \zeta\left(\sigma_{1,2,3};\pi_{1,2}\right)\mathrm{d}\pi_1\mathrm{d}\pi_2 \\ =& 2\int_{\sigma_1}^{+\infty}\int_{\sigma_2}^{+\infty} \frac{r_1 r_2\left[\zeta(r_1,r_2,r_3^+)+\zeta(r_1,r_2,r_3^-)\right]}{\sqrt{(r_1^2-\sigma_1^2)(r_2^2-\sigma_2^2)}}\mathrm{d}r_1\mathrm{d}r_2 ,\end{aligned} \tag{2}$$

in which

$$\begin{aligned}r_3^+ &= \sqrt{\sigma_3^2 + \left(|\pi_1| + |\pi_2|\right)^2} \\ r_3^- &= \sqrt{\sigma_3^2 + \left(|\pi_1| - |\pi_2|\right)^2} .\end{aligned} \tag{3}$$

Szapudi (2004a) pointed out that 3PCF can be expanded with Legendre polynomials P_ℓ to isolate part of the configuration dependence,

$$\begin{aligned}\zeta(r_1,r_2,\theta) &= \sum_{\ell=0}^{\infty}\frac{2\ell+1}{4\pi}\zeta_\ell(r_1,r_2)P_\ell(\cos\theta) \\ \zeta_\ell(r_1,r_2) &= 2\pi\int_{-1}^{1}\zeta(r_1,r_2,\theta)P_\ell(\cos\theta)\mathrm{d}\cos\theta ,\end{aligned} \tag{4}$$

in which $\cos\theta = -\mathbf{r}_1\cdot\mathbf{r}_2/(r_1 r_2)$. In the expansion the monople ζ_0 is of particular interests for its relatively simplicity in measurement and interpretation (Pan & Szapudi 2005a,b). One can easily found that ζ_0 is actually the spherical average of ζ in three-dimensional space

$$\frac{\zeta_0(r_1,r_2)}{4\pi} = \frac{\int\zeta(r_1,r_2,\theta)2\pi\sin\theta\mathrm{d}\theta}{4\pi} = \frac{\int\zeta\mathrm{d}\Omega}{\int\mathrm{d}\Omega} , \tag{5}$$

which effectively becomes theoretical support to the estimator in Pan & Szapudi (2005b).

In the same spirit, the projected 3PCF Z also can be expanded but in a different treatment, the cosine Fourier transformation proposed by Zheng (2004) and Szapudi (2009) is the appropriate one since Z is defined on a two-dimensional plane which is perpendicular to LOS. Angular averaging of Z thus produces the zeroth-order component of the cosine expansion to Z (Zheng 2004; Szapudi 2009),

$$Z_0(\sigma_1,\sigma_2) = \frac{1}{2\pi}\int_0^{2\pi} Z(\sigma_1,\sigma_2,\theta_p)\mathrm{d}\theta_p \tag{6}$$

with $\theta_p = \cos^{-1}[(\sigma_1^2+\sigma_2^2-\sigma_3^2)/(2\sigma_1\sigma_2)]$, which is the object function that we focus on and actually is related to ζ by

$$\begin{aligned}Z_0 =& \frac{1}{2\pi}\int_0^{2\pi}\mathrm{d}\theta_p\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\zeta\left(\sigma_1,\sigma_2,\theta_p;\pi_1,\pi_2\right)\mathrm{d}\pi_1\mathrm{d}\pi_2 \\ =& \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\widetilde{\zeta_0}(\sigma_1,\sigma_2,\pi_1,\pi_2)\mathrm{d}\pi_1\mathrm{d}\pi_2\end{aligned} \tag{7}$$

where

$$\widetilde{\zeta_0}(\sigma_1,\sigma_2,\pi_1,\pi_2) = \frac{1}{2\pi}\int_0^{2\pi}\zeta\left(\sigma_1,\sigma_2,\theta_p;\pi_1,\pi_2\right)\mathrm{d}\theta_p . \tag{8}$$

Note that $\widetilde{\zeta_0}$ and the monople of 3PCF ζ_0 are not equal at all.

Theoretically if the nonlinear bispectrum is known, by the cosine transformation

$$B(k_1,k_2,\phi) = \sum_{n=-\infty}^{+\infty} B_n(k_1,k_2)\cos(n\phi)$$
$$B_n(k_1,k_2) = \frac{1}{2\pi}\int_0^{2\pi} B(k_1,k_2,\phi)\cos(n\phi)\mathrm{d}\phi\,, \tag{9}$$

it is fairly straightforward to compute Z_0 (Zheng 2004)

$$Z_0(\sigma_1,\sigma_2) = \frac{1}{(2\pi)^2}\int_0^{+\infty}\int_0^{+\infty} B_0(k_1,k_2) \times J_0(k_1\sigma_1)J_0(k_2\sigma_2)k_1k_2\mathrm{d}k_1\mathrm{d}k_2\,, \tag{10}$$

where J_0 is the zero-order Bessel function of the first kind. An immediate fact is that Z_0 only requires good approximation to the zeroth-order component of the nonlinear bispectrum, which simplifies theoretical development.

At a first glance it seems that it is not useful to invoke Z_0, since Z contains more information, the former erases the angular dependence completely through averaging. However, by smoothing Z_0 suffers much less from shot-noise than Z, i.e. has smaller variance, which is a celebrated property particularly when sample is not of high number density. More important, as we will see in next section, Z_0 can be easily estimated with the common procedure for anisotropic 2PCF after some minor modification. The savings in computing time, proportional to the number of galaxies, is tremendous compared with calculating the projected 3PCF of the full configuration.

3 ESTIMATION AND NUMERICAL TEST

3.1 Estimator

Estimation of the zeroth-order component of the projected 3PCF is based on Eqs. 7 and 8. Eq. 8 indicates that $\widetilde{\zeta_0}$ can be measured with the same estimator of ζ_0 as in Pan & Szapudi (2005b), taking the same form of the one in Szapudi & Szalay (1998),

$$\widetilde{\zeta_0} = \frac{DDD - 3DDR + 3DRR - RRR}{RRR}\,, \tag{11}$$

grouped symbols of D and R refer to various normalized number counts of triplets similar to what is in Pan & Szapudi (2005b), difference is that $\widetilde{\zeta_0}$ is estimated in bins of both σ and π. Explicitly, if scale bins are linear, given two vector bins $\mathbf{r}_{jk} = (\sigma_j,\pi_k)$ and $\mathbf{r}_{j'k'} = (\sigma_{j'},\pi_{k'})$, with σ_{jk} in $(\sigma_{jk}-\Delta\sigma/2,\sigma_{jk}+\Delta\sigma/2)$, and π_{jk} in $(\pi_{jk}-\Delta\pi/2,\pi_{jk}+\Delta\pi/2)$, as an example, the DDD is obtained through

$$DDD = \begin{cases} \frac{\sum_{i=1}^{N_g} n_i(\mathbf{r}_{jk})\left[n_i(\mathbf{r}_{j'k'})-1\right]}{N_g(N_g-1)(N_g-2)}\,, & \text{if } \mathbf{r}_{jk} = \mathbf{r}_{j'k'} \\ \frac{\sum_{i=1}^{N_g} n_i(\mathbf{r}_{jk})n_i(\mathbf{r}_{j'k'})}{N_g(N_g-1)(N_g-2)}\,, & \text{if } \mathbf{r}_{jk} \neq \mathbf{r}_{j'k'} \end{cases}, \tag{12}$$

where n_i is the number of neighbours to the center point counted in the vector bin $\mathbf{r}_{jk}$. Then by Eq. 7 integrating $\widetilde{\zeta_0}$ over π_k and $\pi_{k'}$ yields estimation of Z_0. We have to address here that unlike 3PCF, the estimator can not completely eliminate edge effects for ζ_0, $\widetilde{\zeta_0}$ and so Z_0, one needs to be cautious when scales at probe is comparable to sample's characteristic size.

3.2 Data preparation and estimation setup

Since our goal is to provide a redshift distortion free third-order statistics, a key question is whether Z_0 measured in redshift space agrees with what we get in real space. In absence of accurate models about redshift distorted 3PCF, particularly in nonlinear regime, the best approach is to work with N-body simulation data directly. Two realizations of LCDM simulations run with Gadget2 (Springel 2005) were analysed. Their cosmological parameters are taken from WMAP3 fits (Spergel et al. 2007), $\Omega_m = 0.236$, $\Omega_\Lambda = 0.764$, $h = 0.73$ and $\sigma_8 = 0.74$. 512^3 particles were evolved in both simulations, but one box size is $L = 300h^{-1}$Mpc (box300) and the other is $L = 600h^{-1}$Mpc (box600), the force softening lengths are $12h^{-1}$kpc and $24h^{-1}$kpc respectively. The $z = 0$ output of box300 simulation and $z = 0.09855$ output of box600 simulation were selected for our numerical experiment.

It is unpractical to use all particles in the simulations therefore for each set of data we generate nine diluted samples for analysis to control the amount of computation at a reasonable level; all results we present here are mean values of nine runs, and the actual scatter of different realizations is very small. For box300 the number of randomly picked points is about 0.2% of the total, while for box600 more than $\sim 600,000$ points are used. Several other samples diluted at different levels were also generated for consistency check. We find that sample dilution does affect our estimation of Z_0 but mainly on very small scales, and that variance due to discreteness becomes larger with fewer points, as expected.

A common assumption about redshift distortions is the plane parallel approximation (distant observer assumption), which assumes that the observer is very far away from the sample so that all lines-of-sight from the observer to galaxies are parallel to each other. It simplifies calculation by reducing a 3D problem into 1D and indeed works well when the interested scale opens only a narrow angle to the observer. But the systematic bias introduced by the plane parallel approximation turns out to be significant if the angle becomes wide. Theoretical calculations and numerical measurements have shown that the deviation mainly occurs at relatively large scales and could be more than 10% (e.g. Szalay et al. 1998; Scoccimarro 2000; Szapudi 2004b; Cai & Pan 2007; Pápai & Szapudi 2008). To test the accuracy of the plane parallel approximation, two sets of samples in redshift space are generated for the box300 data, one set takes the distant observer assumption while the other mimics realistic samples by placing an observer at distance of $100h^{-1}$Mpc to the nearest surface of the sample. One has to bear in mind that the two redshift distortion scenarios differ not only in sample construction but also the way of decomposing the separation $\mathbf{r}$ into (σ,π) amid measurement. The output of box600 simulation we used is of $z \approx 0.1$, the plane parallel approximation is sufficient for most analysis if interested scales are less than $\sim 50h^{-1}$Mpc. We also noticed that applying periodic boundary condition or simply throwing away those points shifted out of box by peculiar velocity makes little difference for the final Z_0, which effectively eliminates the concern of Nock et al. (2010).

During our estimation the σ bins are set logarithmic with $\Delta\log\sigma = \log 1.4$, π bins are linear with $\Delta\pi = 3,5h^{-1}$Mpc for the box300 and the box600 respectively. Caution must be taken about the bin width which shall not be too wide to degrade the accuracy too much, while it shall be large enough to achieve $DDD > 0$ for even the narrowest bin. Experience shows that normally $DDD > 100$ is good enough to give reliable estimation at our accuracy goal of 10 %.

As we do not have multiple realizations to produce error bars, for each simulation data set we split the sample into eight sub-volume boxes in half size, then the scatter of measurements in these eight sub-volume boxes are taken as an estimate of the variance.

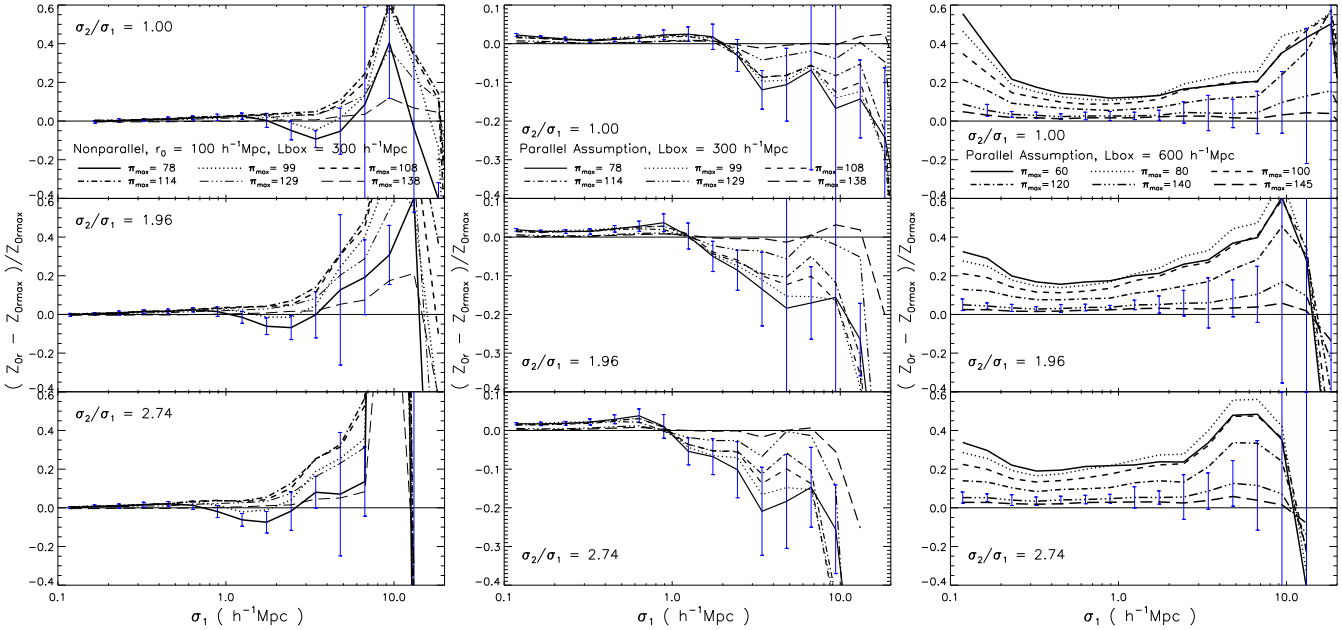

Figure 1. Fractional change of $Z_0(\sigma_1, \sigma_2)$ with π_{max} compared to the one measured with largest π_{max}. Three classes of configuration of σ_2/σ_1 of Z_0 are presented. Z_{0r} measured in real space with π_{max} as specified in the legend, Z_{0rmax} is estimated with the largest $\pi_{max} = 150h^{-1}$Mpc. Plane parallel approximation in real space means merely the decomposition of **r** into (σ, π) using parallel LOS, while nonparallel corresponds to an external observer at the distance of $100h^{-1}$Mpc to the bottom of the simulation box. Errorbars of $\pi_{max} = 78h^{-1}$Mpc for box300, $\pi_{max} = 120h^{-1}$Mpc for box600 are plotted to show the uncertainty.

3.3 Finite integration range along LOS

The integration range along LOS to have Z_0 from $\widetilde{\zeta_0}$ in Eq. (7) should be $(-\infty, +\infty)$ to guarantee conservation of triplets along LOS. However, one can not integrate infinite scales due to finite radial thickness of realistic samples, so there is always a finite upper limit of π to the integration. Our measurements thus correspond to

$$\hat{Z}_0 = \int_{-\pi_{max}}^{+\pi_{max}} \int_{-\pi_{max}}^{+\pi_{max}} \widetilde{\zeta_0} \mathrm{d}\pi_1 \mathrm{d}\pi_2 = \sum_{i,j} \widetilde{\zeta_0} \Delta\pi_i \Delta\pi_j \ . \quad (13)$$

Let subscript r denote quantities in real space and s for those in redshift space. The practical limitation certainly introduces systematic bias, henceforth mathematically $\hat{Z}_{0s} \neq \hat{Z}_{0r} \neq Z_0$. What we hope is that we can find a π_{max} so that the contribution from π larger than that is negligible at a given tolerance. In our test runs we found that the largest π_{max} permitted is around $1/4 - 1/2$ of the box size. If larger scale is used, the estimator of Z_0 suffers greatly of finite-volume effects. The same problem is present when estimating projected 2PCF, and normally it is agreed that $\pi_{max} \sim 40 - 70h^{-1}$Mpc is sufficient to give stable results at small σ of less than $\sim 20 - 30h^{-1}$Mpc, but may not be enough for measurement on larger scales (see Baldauf et al. 2010, and references there in).

Figure 1 presents the convergence of measurements with changing π_{max}. It displays the fractional differences of Z_0 compared to that calculated from the largest π_{max} allowed by the geometry of sample. Samples used in this test are all in real space, but for box300 data we decompose scales by LOS in two ways: plane parallel approximation and wide angle treatment.

For box300 at scales $\sigma < 1h^{-1}$Mpc Z_0 is extremely stable against different choices of π_{max}, independent of the scheme of scale decomposition, but at larger σ scales the influence of those large π becomes more and more evident. It appears that in the wide angle treatment Z_0 actually increases with π_{max} up to $\sim 110h^{-1}$Mpc and then falls down when further enlarging π_{max}, while in the plane parallel assumption Z_0 monotonically rises with larger π_{max}. The results from box600 are somewhat different: Z_0 decreases with increasing π_{max} on all σ scales. Additional numerical experiments with the box600 data revealed that this behaviour is largely caused by the dilution of the original data: a denser sample has less variation against the choice of π_{max} when $\sigma < 1h^{-1}$Mpc.

We conclude that for an overall precision target of $\sim 10\%$ for σ scales below $10h^{-1}$Mpc, $\pi_{max} \approx 120h^{-1}$Mpc suffices. This is much larger than customary for the projected 2PCF. Note that the sharp break down of convergence at scales $\sigma \sim 10 - 20h^{-1}$Mpc appears to be a numerical artifacts where Z_0 quickly approaches zero.

3.4 Redshift space *versus* real space

Figure 2 demonstrates Z_0 of the two simulation data sets in redshift space and real space for different σ_2/σ_1 and π_{max}; detailed comparison is drawn in Figure 3. On most scales of $\sigma < 10h^{-1}$Mpc, residual effects of redshift distortion due to finite integration domain result in only a minor bias within 10%, except for an upshot in Z_0 in redshift space on scales of $\sigma <\sim 0.2h^{-1}$Mpc. Adjusting π_{max} does not modulate Z_0 significantly on $\sigma <\sim 1h^{-1}$Mpc, but causes some apparent deviations on larger scales, especially where Z_0 approaches its zero-crossing point. Nevertheless, it is reassuring from Figures 2 and 3 that Z_0 estimated in redshift space agrees well with that of real space with at most 10% uncertainty for $0.2 < \sigma <\sim 10h^{-1}$Mpc and $\pi_{max} \sim 120h^{-1}$Mpc. Thus Z_0 can be accepted as a redshift distortion free third order statistics to a good precision.

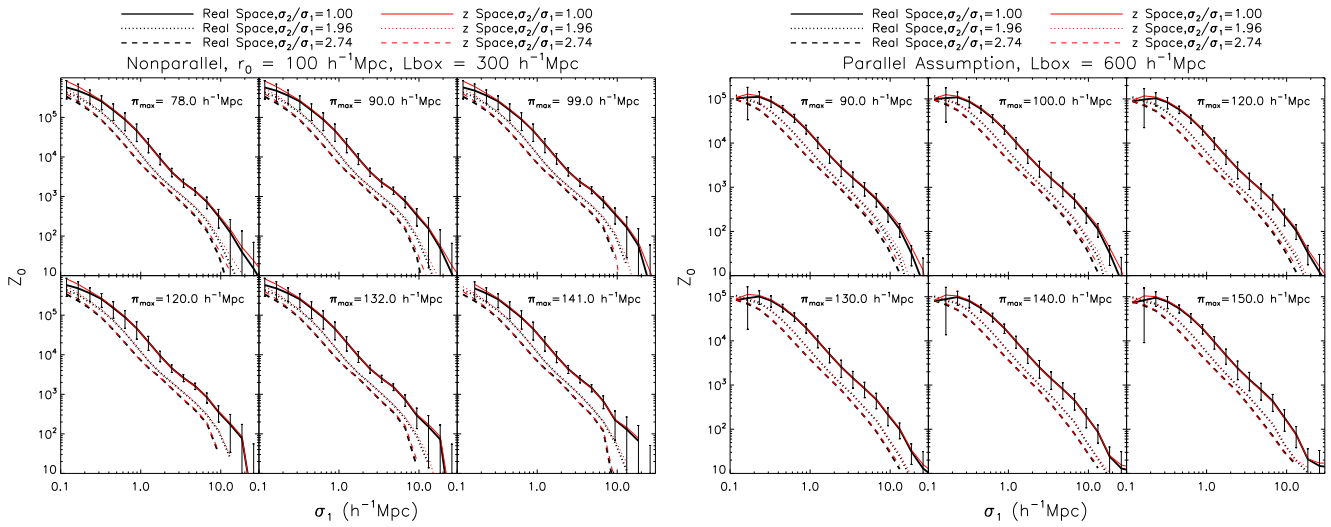

Figure 2. Z_0 measured in real space (thick lines) and that in redshift space (thin lines) for different σ_2/σ_1 and π_{max}. Errorbars are of measurements of configuration $\sigma_2/\sigma_1 = 1$ in real space. The results not shown for box300 under plane parallel approximation are similar to the non-parallel case.

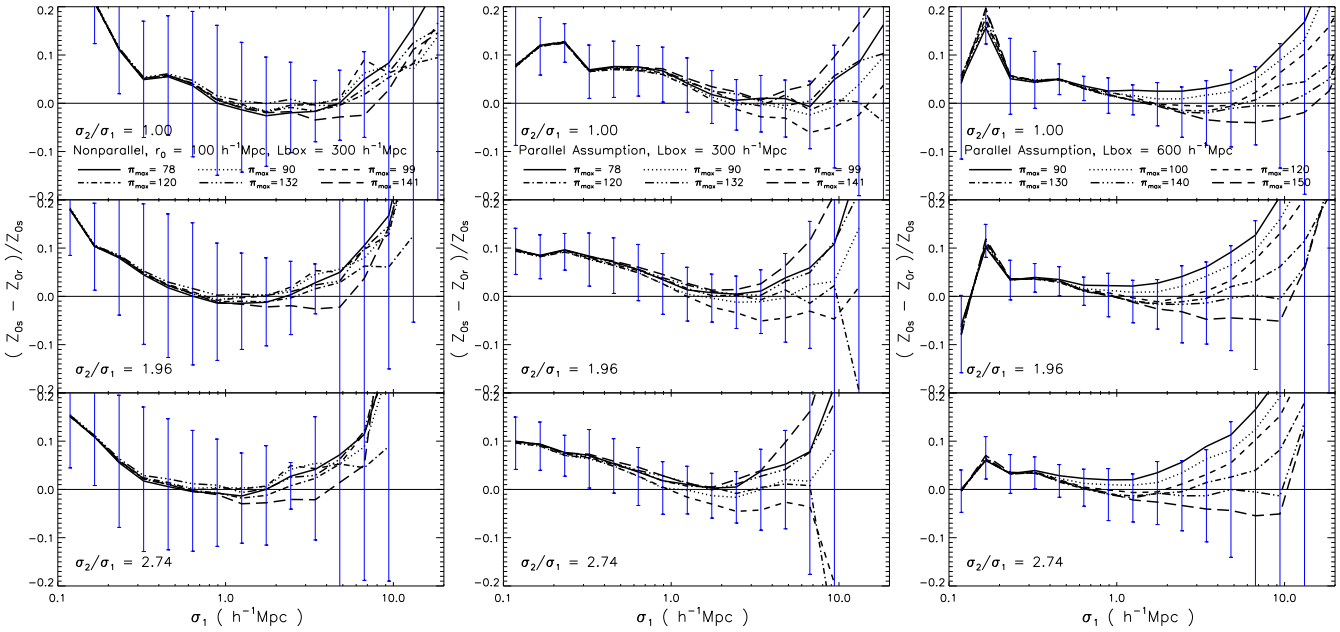

Figure 3. Deviation of Z_0 in redshift space to Z_0 in real space for different configurations of σ_2/σ_1 and choices of π_{max}. Left panel shows the wide angle treatment to the redshift distortion to the box300 data, the middle shows the box300 results under plane parallel assumption, and the right panel is shows box600 with plane parallel redshift distortion.

4 HALO MODEL PREDICTION OF Z_0

4.1 Formalism

The halo model invoked to model the third-order statistics Z_0 of dark matter basically follows Ma & Fry (2000a,b), Scoccimarro et al. (2001), Fosalba et al. (2005) and Smith et al. (2008). Here we just give a brief description of main ingredients of the model, for more details we refer to the review of Cooray & Sheth (2002).

(i) Halo profile $\rho(r)$. It has been pointed out that the density profile of a virialized dark matter halo in general is ellipsoidal and shows various morphology rather than a simple universal spherical approximation (Jing & Suto 2000, 2002). The non-spherical shape of halo can evidently affect the halo model prediction of the clustering of dark matter on small scales where the one-halo term dominates (Smith et al. 2006). Noting that Z_0 is a degenerated 3PCF in analogous to ζ_0, and should be similarly insensitive to halo shapes, the popular NFW profile (Navarro et al. 1997) is still adequate for our model. For a halo of mass M it reads,

$$u(r) = \frac{\rho(r)}{M} = \frac{fc^3}{4\pi R_v^3} \frac{1}{cr/R_v(1+cr/R_v)^2}\,, \qquad (14)$$

where $f = [\ln(1+c) - c/(1+c)]^{-1}$ and $c(M) =$

$c_0(M/M_*)^{-\beta}$ known as the concentration parameter; parameters $c_0 = 9$, $\beta = 0.13$ are calibrated by numerical simulation (Bullock et al. 2001). Halo mass is defined as $M = (4\pi R_v^3/3)\Delta\rho_{crit}$ with ρ_{crit} being the cosmological critical density. Δ is the density contrast for virialization and can be estimated from spherical collapse model. A good fit for a flat universe with cosmological constant is given by Bryan & Norman (1998)

$$\Delta = \frac{18\pi^2 + 83x - 39x^2}{\Omega(a)} \ , \ x = \Omega(a) - 1 \ . \tag{15}$$

As it is more convenient to work in Fourier space for 3PCF, the Fourier transformed halo profile of Scoccimarro et al. (2001) is the used for computations

$$\begin{aligned} u(k, M) = & f\Big\{ \sin\eta \left[\mathrm{Si}(\eta[1+\eta]) - \mathrm{Si}(\eta)\right] \\ & + \cos\eta \left[\mathrm{Ci}(\eta[1+\eta]) - \mathrm{Ci}(\eta)\right] - \frac{\sin\eta c}{\eta(1+c)} \Big\} \end{aligned} \tag{16}$$

with $\eta = kR_v/c$. Note that the halo profile is presumably truncated at virial radius R_v in the standard version of halo model for large scale structure; in extensions it becomes an adjustable parameter in an attempt to find the best match to simulations.

(ii) Mass function $n(M)$. There are many versions of halo mass functions, but it turns out that using mass function with higher precision actually brings about only a relatively minor change to Z_0 on the small scales as we tested. The classical Sheth-Tormen function (Sheth & Tormen 1999) is sufficient,

$$n(M)M\mathrm{d}M = \bar{\rho}\frac{dy}{y}A\gamma\sqrt{\frac{g(\nu)}{2\pi}}\left(1 + g(\nu)^{-p}\right)\exp(-\frac{g(\nu)}{2}) \tag{17}$$

in which $\gamma = \mathrm{d}\ln\sigma_M^2/\mathrm{d}\ln R$, $g(\nu) = \alpha\nu^2$, $\nu = \delta_c/\sigma_M$, $y = (M/M_*)^{1/3} = R/R_*$ with $R_* = R_v\Delta$ and σ_M being the extrapolated linear variance of the dark matter fluctuations smoothed over the Lagrangian scale R. $A = 0.322$ $\alpha = 0.707$, $p = 0.3$ are parameters fitted to simulations by Jenkins et al. (2001).

Note that the definition of halo mass in the Sheth-Tormen function is $M = 4\pi\bar{\rho}\Delta R_v^3/3$ with $\bar{\rho} = 2.78 \times 10^{11}\Omega_m h^2 M_\odot \mathrm{Mpc}^{-3}$ being the dark matter density of the present universe, while the halo mass in in the NFW profile is defined with the critical mass $\rho_{crit} = \bar{\rho}/\Omega_m$, conversion between halo parameters of the two sets is given in Smith & Watts (2005). Scoccimarro et al. (2001) already noticed the inconsistency but argued that effects of the difference could be largely cancelled in practical calculations, and Fosalba et al. (2005) also find that changing the concentration parameter by as much as 50% would not affect the final results significantly. This is also the case in our calculation.

(iii) Halo bias. The distribution of massive halos is biased to the dark matter. Most halo bias functions extracted from N-body simulations (e.g. Mo et al. 1997; Jing 1999; Sheth & Tormen 1999; Tinker et al. 2010) are refinements to the analytical model of Mo & White (1996). The bias plan used in our recipe is the fitting formula given by Sheth & Tormen (1999),

$$b(\nu) = 1 + \frac{g(\nu) - 1}{\delta_c} + \frac{2p}{\delta_c(1 + g(\nu)^p)} \ , \tag{18}$$

in which δ_c is the linear overdensity threshold for spherical collapse. Its cosmological dependence is so weak that a constant value of 1.686 is usually taken. The most recent update of Tinker et al. (2010) is also applied in our code for a consistency check, and the results indicate that the improvement to Z_0 is minor in the intermediate nonlinear regime only; it does not bring significant improvement to the overall accuracy when considering the magnitude of numerical errors of estimation of the previous section.

To prevent multi-dimensional integrations involved in direct calculation of ζ in configuration space (Takada & Jain 2003), we work in Fourier space to yield the bispectrum B predicted by halo model first. Then B_0 is easily obtained to render Z_0 through the transformation of Eq. 10. In the halo model, bispectrum consists of three separate terms, namely the one-halo, two-halo and three-halo terms,

$$B(k_1, k_2, \phi) = B_{1h} + B_{2h} + B_{3h} \ , \tag{19}$$

in which

$$\begin{aligned} B_{1h} &= I_{03}(k_1, k_2, k_3) \\ B_{2h} &= I_{11}(k_1)I_{12}(k_2, k_3)P_L(k_1) + cyc. \\ B_{3h} &= \left[\prod_{i=1}^{3} I_{11}(k_i)\right] B_{PT} \end{aligned} \tag{20}$$

and

$$I_{ij} = \int \frac{dr}{r} n(r) b_i(r) [u(k_1, r) \dots u(k_j, r)] \left(\frac{4\pi r^3}{3}\right)^{j-1} \tag{21}$$

with $b_0 = 1$, $b_1 = b(\nu)$ and $b_i = 0$ for $i > 1$ to neglect quadratic and high order biasing terms. P_L is the linear power spectrum and it is generated by CMBFAST (Seljak & Zaldarriaga 1996) with the cosmological parameters from the simulations we use. B_{PT} is the bispectrum predicted by the Eulerian perturbation theory at tree level (e.g. Bernardeau et al. 2002).

4.2 Comparison with simulations

Z_0 predicted by halo model and Eulerian perturbation theory is demonstrated in Figures 4 and 5 overlaid with measurements of the box300 and the box600 simulation data, respectively, estimated in real space. Results of our halo model and simulations agree remarkably well at both redshifts $z = 0, 0.1$, especially on σ_1 scales between $\sim 0.2-$ $\sim 5h^{-1}$Mpc.

On very small scales $\sigma_1 <\sim 0.2h^{-1}$Mpc, the halo model predicted Z_0 is larger and steeper than the simulation. This is more apparent for box600. Numerical tests reveal that this is partly due to dilution to the original data set: a higher density of points leads a higher clustering power in this regime. On large scales, where halo model follows perturbation theory, both theories begin to overpredict the clustering strength of simulations for larger σ_2/σ_1, which should not be attributed to the imperfection of halo models and should be the inaccuracy of B_{PT} on these scales (Pan et al. 2007; Guo & Jing 2009).

To improve halo model performance at the three-point level, halo boundary and mass function adjustments are usually adopted (Takada & Jain 2003; Wang et al. 2004; Fosalba et al. 2005). This alleviates the disagreement to some extent. Here we also enlarge halo boundary beyond R_v and truncate the high mass tail of halo mass function (Figures 4 and 5). Experiment indicates that this extension of halo radius without a hard cut-off of the mass function can easily generate the correct shape and amplitude of Z_0 of simulations. Simple fitting shows that best halo boundary is $\sim 1.5R_v$ for box300 and $\sim 1.6R_v$ for box600. In contrast, if we keep the halo boundary unchanged but truncate the halo mass function, the one- and two-halo terms are so strongly modified, and the shape and the amplitude of Z_0 deviate from simulations significantly. Simple fitting to simulations by setting both halo boundary and mass cut-off

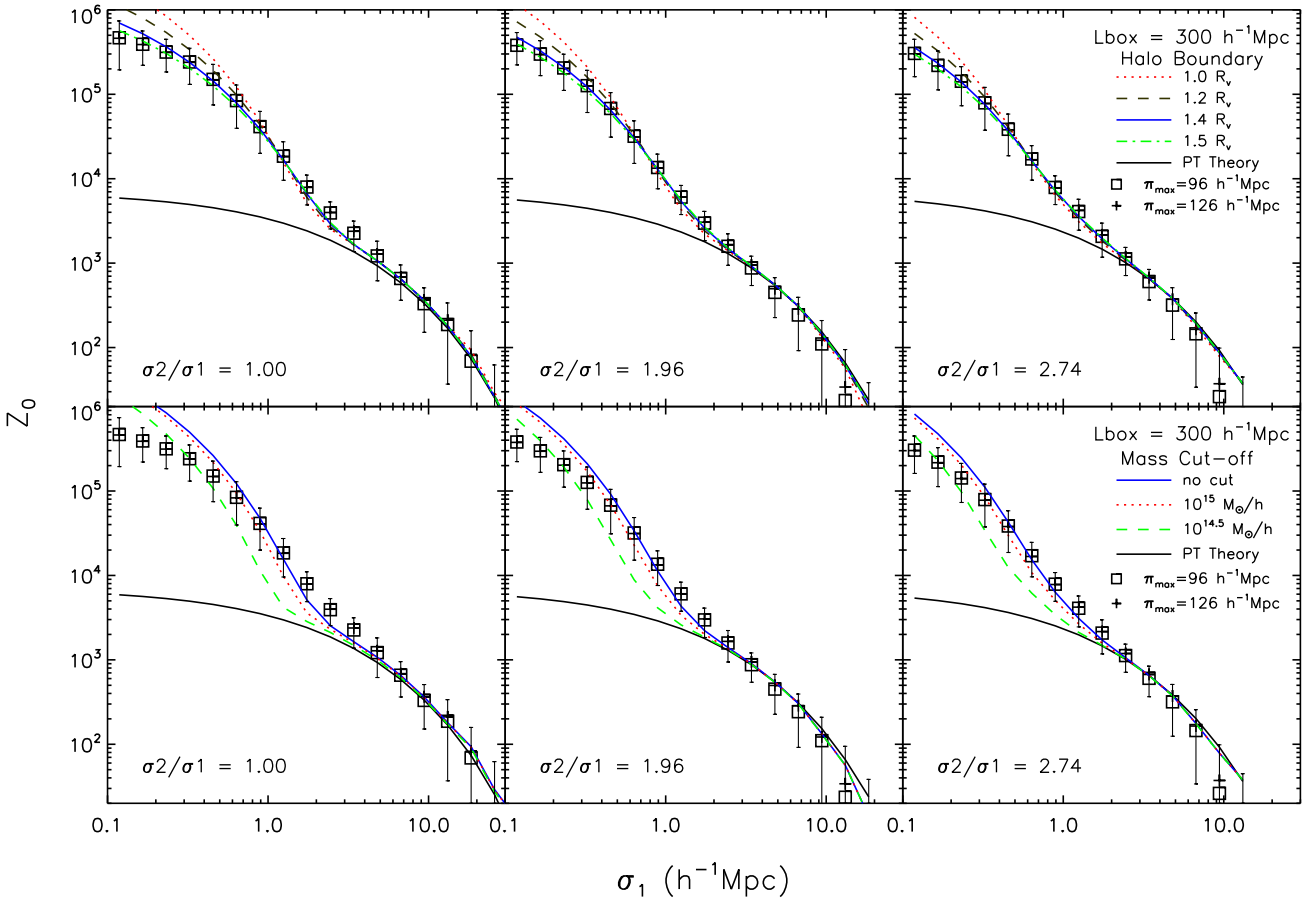

Figure 4. Z_0 of the box300 simulation (z=0) with predictions from halo model and Eulerian perturbation theory (PT). Symbols are measurements of simulations in real space with different π_{max}. The upper row of plots shows the effects of adjusting halo boundary radius in unit of R_v but without a cut-off to the halo mass function, while the bottom row demonstrates the consequence of cutting the high mass tails of the mass function with the halo boundary fixed to R_v.

free reveals that mass cut-off could not be smaller than $10^{15}M_{\odot}$. Otherwise, there is no way to reconcile the under-predicted Z_0 with simulations at transition scales of $\sigma_1 \sim 3h^{-1}$Mpc, above which B_{PT} breaks. In conclusion, enlarging halo boundary alone is sufficient for for accurately predicting Z_0.

During our calculation, we also examine the influence of the halo bias function and the mass function by using the high precision formula of Tinker et al. (2010). Such replacement does not cause a fundamental change to the theoretical prediction (Figure 6). The new mass functions do not benefit the halo model much. On most σ scales, less than 10^{-1}Mpc, the halo model with Sheth-Tormen functions is consistent with simulations within our error budget of $\sim 10-20\%$. The replacement of functions provided by Tinker et al. (2010) increases deviation level to around $20-30\%$, especially on scales of $\sim 1h^{-1}$Mpc; visible advantage only just appears on scales of $\sigma_1 >\sim 3h^{-1}$Mpc with accuracy gain of a few percents.

In addition to the halo model, we also checked the phenomenological models of Scoccimarro & Couchman (2001) and Pan et al. (2007). The accuracy of the formula by Scoccimarro & Couchman (2001) is very good on scales $\sigma_1 >\sim 1h^{-1}$Mpc but then deviates from the simulations by more than 40% on smaller scales. The performance of Pan et al. (2007) is poor in terms of Z_0 as the bispectrum model is not designed to conserve clustering power and the resulting integration over it yields incorrect amplitude. Nevertheless, if a renormalization is enforced for the model to be consistent with the perturbation theory on large σ, the model works well for Z_0 at $\sigma_1 >\sim 2h^{-1}$Mpc.

5 SUMMARY AND DISCUSSION

In this paper we propose a third-order correlation function for characterising galaxy clustering properties. The statistics Z_0 we advocate is the zeroth-order component of the projected 3PCF. Although Z_0 is a 3PCF, its estimation takes roughly the same amount of computing operation as the projected 2PCF. The algorithm can be eas-

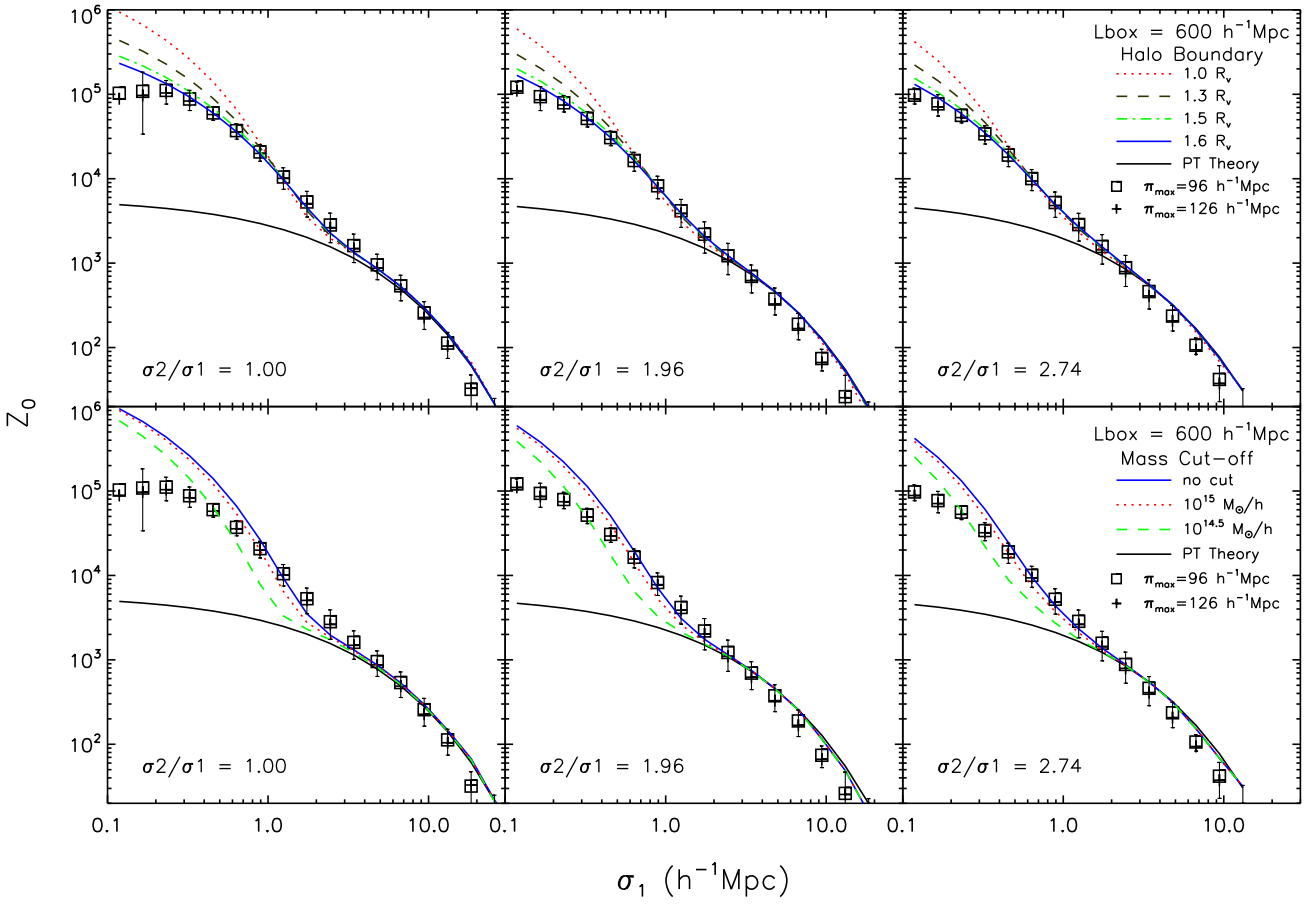

Figure 5. Z_0 of the box600 simulation data and models. See last figure for details.

ily implemented after moderate modification of a code for the projected 2PCF.

Various numerical experiments confirm that Z_0 can be deemed to be redshift distortion free within approximately 10% for the regime where the scale perpendicular to LOS is $0.2 < \sigma < 10h^{-1}$Mpc. In addition, the maximal integration scale π_{max} parallel to LOS during estimation ought to be greater than $\sim 120h^{-1}$Mpc. A serious concern is that shot noise could ruin the estimation in the strongly nonlinear regime if the number density of points in a sample is too low. This requirement for a robust Z_0 measurement is tighter than for the projected 2PCF, but still weaker than the normal projected 3PCF, since Z_0 is an integral of the former. The criterion we suggest is $DDD >\sim 100$.

As we expected, the halo model provides satisfactory prediction to dark matter Z_0 of simulations within $\sim 10\%$, if the classical Sheth-Tormen mass functions are used. Our computation indicates that extending the halo boundary is enough to yield good fit to simulations, while a hard cut-off to mass function is not as effective as previous works claimed. Substituting new functions of the halo mass distribution and halo biasing in high precision does not lead to significantly better agreement with simulations. Since the angular dependence in the projected 3PCF and the normal 3PCF is smeared out in Z_0, we conjecture that using an anisotropic halo profile probably will not significantly improve accuracy. A significant bias of halo model predicted Z_0 compared to simulations emerges in the weakly nonlinear regime, where halo models boil down to second-order perturbation theory; the latter is already known to be poor in predicting dark matter 3PCF. A more precise bispectrum from higher order perturbation theories may offer a way to increase precision (e.g. Valageas 2008; Sefusatti 2009; Bartolo et al. 2010).

The principal reason for proposing Z_0 is to provide an efficient redshift distortion free 3PCF, complementary to the standard projected 2PCF, for galaxy clustering analyses. It is well known that the projected 2PCF itself is a Gaussian statistic only and thus has its limitations. Third order correlation functions, mainly carrying information about non-Gaussianity, are more sensitive to details of the galaxy distribution. Non-Gaussianity of galaxy distribution is generated by the nonlinear action of gravitational force and gas physics if the primordial density fluctuation of the universe after inflation is Gaussian. The degeneracy shown in projected 2PCF (e.g. Zu et al. 2008) may be broken if third order correlation functions are employed. The redshift distortion free feature of Z_0 on scales

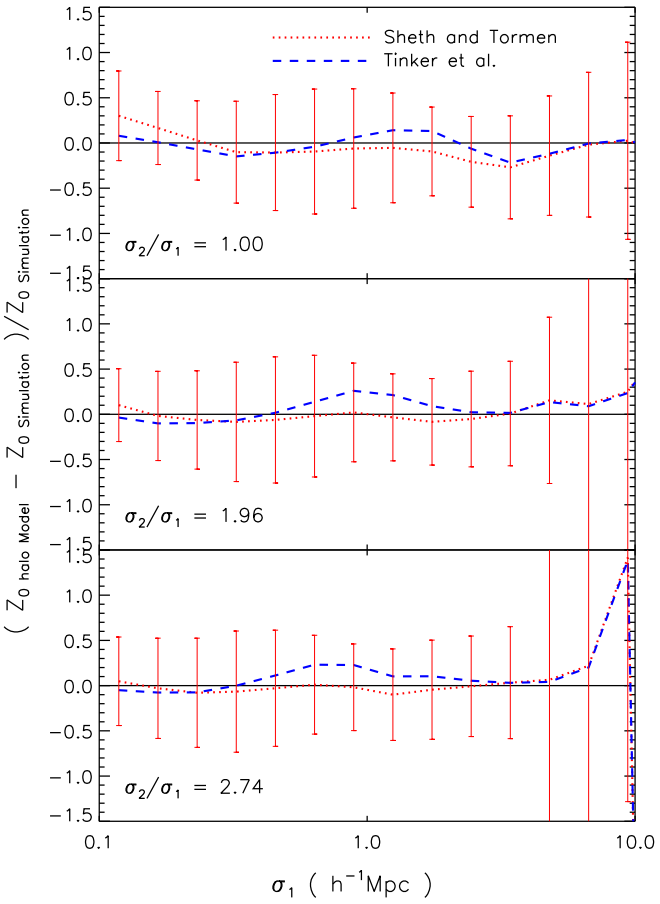

Figure 6. Relative differences of the halo model predicted Z_0 to that of estimated from the box300 simulation. Red dotted lines correspond to the prediction of the halo model with halo mass function and bias function given by Sheth & Tormen (1999), blue dashed lines are generated by the model with mass function and bias function provided by Tinker et al. (2010). $Z_{0\,simulation}$ is the estimation from the box300 data with plane-parallel assumption to the redshift distortions.

less than $10h^{-1}$Mpc defines its potential in investigating the relation of galaxies with their host halos, and the formation histories of galaxies and halos. Furthermore, the success of halo model prediction on dark matter Z_0 encourages us to apply Z_0 for analysing galaxies. In principle, with measurements from galaxy samples, Z_0 enables us to generalize and diagnose schemes of HOD, conditional luminosity function (CLF, Yang et al. 2003) and semi-analytical models (e.g. Baugh 2006) to third order statistics at cost of one additional free parameter, the halo boundary. Our present work is restricted to dark matter only, the behavior of Z_0 for biased objects remains unclear. Testing with mock galaxy samples before applying to real data will be necessary.

ACKNOWLEDGMENT

This work is supported by the Ministry of Science & Technology of China through 973 grant of No. 2007CB815402 and the NSFC through grants of Nos. 10633040, 10873035. JP acknowledges the One-Hundred-Talent fellowship of CAS. IS acknowledges support from NASA grants NNG06GE71G and NNX10AD53G and from the Polányi Program of the Hungarian National Office for Research and Technology (NKTH). We thank Weipeng Lin for his kindness of providing us N-body simulation data.

REFERENCES

Baldauf T., Smith R. E., Seljak U., Mandelbaum R., 2010, Phys. Rev. D, 81, 063531
Bartolo N., Beltrán Almeida J. P., Matarrese S., Pietroni M., Riotto A., 2010, Journal of Cosmology and Astro-Particle Physics, 3, 11
Baugh C. M., 2006, Reports on Progress in Physics, 69, 3101
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, Phys. Rep., 367, 1
Bryan G. L., Norman M. L., 1998, ApJ, 495, 80
Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001, MNRAS, 321, 559
Cai Y., Pan J., 2007, Chinese Journal of Astronomy and Astrophysics, 7, 51
Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
Davis M., Peebles P. J. E., 1983, ApJ, 267, 465
Fosalba P., Pan J., Szapudi I., 2005, ApJ, 632, 29
Guo H., Jing Y. P., 2009, ApJ, 698, 479
Hawkins E., et al., 2003, MNRAS, 346, 78
Hivon E., Bouchet F. R., Colombi S., Juszkiewicz R., 1995, A&A, 298, 643
Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, MNRAS, 321, 372
Jeong D., Komatsu E., 2009, ApJ, 703, 1230
Jing Y. P., 1999, ApJ, 515, L45
Jing Y. P., Börner G., 1998, ApJ, 503, 37
Jing Y. P., Börner G., 2004, ApJ, 607, 140
Jing Y. P., Suto Y., 2000, ApJ, 529, L69
Jing Y. P., Suto Y., 2002, ApJ, 574, 538
Kang X., Jing Y. P., Mo H. J., Börner G., 2002, MNRAS, 336, 892
Kayo I., et al., 2004, PASJ, 56, 415
Ma C., Fry J. N., 2000a, ApJ, 543, 503
Ma C., Fry J. N., 2000b, ApJ, 531, L87
Matsubara T., Suto Y., 1994, ApJ, 420, 497
McBride C. K., Connolly A. J., Gardner J. P., Scranton R., Newman J. A., Scoccimarro R., Zehavi I., Schneider D. P., 2010, ArXiv e-prints
Mo H. J., Jing Y. P., White S. D. M., 1997, MNRAS, 284, 189
Mo H. J., White S. D. M., 1996, MNRAS, 282, 347
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Nichol R. C., et al., 2006, MNRAS, 368, 1507
Nock K., Percival W. J., Ross A. J., 2010, MNRAS, 407, 520
Pan J., Coles P., Szapudi I., 2007, MNRAS, 382, 1460
Pan J., Szapudi I., 2005a, MNRAS, 361, 357
Pan J., Szapudi I., 2005b, MNRAS, 362, 1363
Pápai P., Szapudi I., 2008, MNRAS, 389, 292
Peebles P. J. E., 1980, The large-scale structure of the universe. Princeton, N.J., Princeton University Press
Scoccimarro R., 2000, ApJ, 544, 597
Scoccimarro R., 2004, Phys. Rev. D, 70, 083007
Scoccimarro R., Couchman H. M. P., 2001, MNRAS, 325, 1312
Scoccimarro R., Couchman H. M. P., Frieman J. A., 1999, ApJ, 517, 531

Scoccimarro R., Sheth R. K., Hui L., Jain B., 2001, ApJ, 546, 20
Sefusatti E., 2009, Phys. Rev. D, 80, 123002
Sefusatti E., Crocce M., Desjacques V., 2010, MNRAS, 406, 1014
Seljak U., 2001, MNRAS, 325, 1359
Seljak U., Zaldarriaga M., 1996, ApJ, 469, 437
Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
Smith R. E., Sheth R. K., Scoccimarro R., 2008, Phys. Rev. D, 78, 023523
Smith R. E., Watts P. I. R., 2005, MNRAS, 360, 203
Smith R. E., Watts P. I. R., Sheth R. K., 2006, MNRAS, 365, 214
Spergel D. N., et al., 2007, ApJS, 170, 377
Springel V., 2005, MNRAS, 364, 1105
Szalay A. S., Matsubara T., Landy S. D., 1998, ApJ, 498, L1
Szapudi I., 2004a, ApJ, 605, L89
Szapudi I., 2004b, ApJ, 614, 51
Szapudi I., 2009, in V. J. Martínez, E. Saar, E. Martínez-González, & M.-J. Pons-Bordería ed., Data Analysis in Cosmology Vol. 665 of Lecture Notes in Physics, Berlin Springer Verlag, Introduction to Higher Order Spatial Statistics in Cosmology. pp 457–492, astro-ph/0505391
Szapudi S., Szalay A. S., 1998, ApJ, 494, L41
Takada M., Jain B., 2003, MNRAS, 340, 580
Tinker J. L., 2007, MNRAS, 374, 477
Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, ApJ, 724, 878
Valageas P., 2008, A&A, 484, 79
Verde L., Heavens A. F., Matarrese S., Moscardini L., 1998, MNRAS, 300, 747
Wang Y., Yang X., Mo H. J., van den Bosch F. C., Chu Y., 2004, MNRAS, 353, 287
White M., 2001, MNRAS, 321, 1
Yang X., Mo H. J., van den Bosch F. C., 2003, MNRAS, 339, 1057
Zehavi I., et al., 2010, astro-ph.CO/1005.2413
Zheng Z., 2004, ApJ, 614, 527
Zheng Z., Weinberg D. H., 2007, ApJ, 659, 1
Zu Y., Zheng Z., Zhu G., Jing Y. P., 2008, ApJ, 686, 41